Forensic Science International: Genetics 5 (2011) 269–275

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems

Carolyn R. Hill [a,*], David L. Duewer [a], Margaret C. Kline [a], Cynthia J. Sprecher [b], Robert S. McLaren [b], Dawn R. Rabbach [b], Benjamin E. Krenke [b], Martin G. Ensenberger [b], Patricia M. Fulmer [b], Douglas R. Storts [b], John M. Butler [a]

[a] National Institute of Standards and Technology, Chemical Science and Technology Laboratory, Gaithersburg, MD 20899-8312, USA
[b] Promega Corporation, Madison, WI 53711-5399, USA

ARTICLE INFO

Article history:
Received 16 February 2010
Received in revised form 19 March 2010
Accepted 23 March 2010

Keywords:
Short tandem repeat
DNA typing
STR kits
Concordance
Multiplex PCR
Population data
D1S1656
D2S441
D2S1338
D3S1358
D8S1179
D10S1248
D12S391
D16S539
D18S51
D19S433
D21S11
D22S1045
FGA
TH01
vWA
SE33
Amelogenin

ABSTRACT

The PowerPlex® ESX 17 and ESI 17 Systems for short tandem repeat (STR) amplification were developed by the Promega Corporation to meet the European Network of Forensic Science Institutes (ENFSI) and the European DNA Profiling (EDNAP) Group recommendations for increasing the number of STR loci included in the European Standard Set (ESS). The PowerPlex ESX 17 and ESI 17 Systems utilize different PCR primer combinations to co-amplify the following 17 loci: D1S1656, D2S441, D2S1338, D3S1358, D8S1179, D10S1248, D12S391, D16S539, D18S51, D19S433, D21S11, D22S1045, FGA, TH01, vWA, SE33, and the sex-typing locus amelogenin. A total of 1443 U.S. population samples were evaluated with pre-commercialization versions of both kits. Stutter and heterozygote peak height ratios have been used to characterize kit performance. Typing results have been used to estimate the match probabilities provided by the chosen loci as well as in concordance studies. Full concordance between the typing results for the two kits was observed in 99.994% (49,055 out of 49,062) STR allele calls compared. All genotyping discrepancies were confirmed by DNA sequence analysis. As a result of these comparisons, a second forward primer for the D22S1045 locus has been added to the PowerPlex ESX 17 System to address a primer binding site mutation and the D1S1656 locus reverse primer in the PowerPlex ESI 17 System was modified to eliminate an amplification-efficiency reducing primer dimer.

Published by Elsevier Ireland Ltd.

1. Introduction

As many national DNA databases are growing at a rapid rate, concern for the potential of false matches with a large number of comparisons being made within and between databases has prompted the desire to add additional loci [1–3]. This is particularly important in Europe where cross-border criminal investigations need to be able to compare DNA data between countries [3,4]. The original European Standard Set (ESS) includes only seven loci: D3S1358, D8S1179, D18S51, D21S11, FGA, TH01, and vWA [3]. In 2006, the European Network of Forensic Science Institutes (ENFSI) and the European DNA Profiling (EDNAP) groups published recommendations to extend the ESS loci [5,6] by adopting three miniSTR loci [7]: D2S441, D10S1248, and D22S1045, as well as two additional polymorphic loci [8]: D1S1656 and D12S391.

The recently released PowerPlex® European Systems from the Promega Corporation (Madison, WI) were created to include these

* Corresponding author. Tel.: +1 301 975 4275.
E-mail address: becky.hill@nist.gov (C.R. Hill).

1872-4973/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.fsigen.2010.03.014

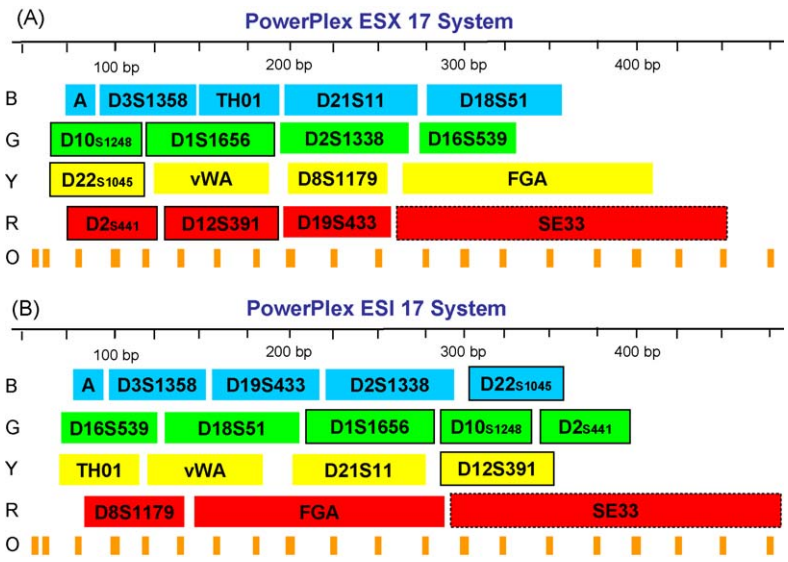

Fig. 1. Schematic of PCR product size ranges and dye color configurations for the STR loci present in (A) PP-ESX17 and (B) PP-ESI17. The "A" in the blue channel is amelogenin. B = blue channel, G = green channel, Y = yellow channel, R = red channel, O = orange channel. The ILS 500 is represented in the orange channel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

extra loci and meet the ENFSI and EDNAP 2006 requests [9]. The PowerPlex ESX 17 (PP-ESX17) and PowerPlex ESI 17 (PP-ESI17) Systems allow co-amplification of 17 STR loci with different size range and dye label configurations as shown in Fig. 1 [10,11]. The PowerPlex ESX 16 (PP-ESX16) and PowerPlex ESI 16 (PP-ESI16) Systems are the same as the PP-ESX17 and -ESI17 Systems with the exception that they do not co-amplify the SE33 locus.

While the PP-ESX17 and PP-ESI17 Systems amplify the same STR loci, different polymerase chain reaction (PCR) primer sequences are utilized for 13 of the 16 STR loci (including SE33) enabling cross-checking for potential primer binding site mutations when the same DNA samples are examined with both kits [12,13]. In order to assess the frequency of allele dropout due to primer binding site mutations, concordance studies were performed between the two new kits as well as other widely used kits [14,15]. In addition, heterozygote peak height ratios and stutter percentages were evaluated as part of characterizing kit performance [16]. Population variation for the 16 STR loci in U.S. Caucasian, African American, Hispanic, and Asian groups are also reported [17].

2. Materials and methods[1]

2.1. DNA samples

Anonymous liquid blood samples with self-identified ethnicities were purchased from Interstate Blood Bank (Memphis, TN) and Millennium Biotech, Inc. (Ft. Lauderdale, FL) and extracted, quantified, and previously typed with the Identifiler® kit (Applied Biosystems, Foster City, CA) [18], the PowerPlex® 16 (PP16) System (Promega, Madison, WI) (data not published), MiniFiler® kit (Applied Biosystems) [19], and the three in-house assays: NIST-23plex, -NC01 and -NC02 [20]. A set of father and son samples previously used for mutation rate studies were also evaluated [21]. The 10 genomic components of Standard Reference Material (SRM) 2391b PCR-based DNA Profiling Standard [22], K562 (Promega), and ABI 007 (Applied Biosystems) were tested for concordance to certified materials and common positive controls. A total of 1461 samples were evaluated in this study with 1455 providing complete genotyping data for concordance testing and 1443 used for peak height ratio and stutter calculations (Table 1).

2.2. PCR amplification and detection

Prototype versions of the PP-ESX17 and PP-ESI17 Systems were used in this study. With the exception of two minor primer changes in D22S1045 and D1S1656, the evaluated materials have the same PCR primers as the now released commercial kits. An additional D22S1045 forward primer was added to correct for a primer binding site mutation in the PP-ESX17 System. The D1S1656 reverse primer in the PP-ESI17 System was changed to avoid generating a primer dimer with the D21S11 labeled primer that reduced amplification at both loci (D21S11 and D1S1656).

These STR kits allow a single amplification of 17 loci in a five-color detection platform using four channels for the PCR products and the fifth channel for the size standard (Fig. 1). All STR assays were run in accordance with manufacturer's recommendations. Each PCR reaction contained 5 μL PP-ESX or PP-ESI 5X Master Mix, 2.5 μL PP-ESX17 or PP-ESI17 10X Primer Set, 16.5 μL diH_2O, and 1 μL of DNA template (0.5–1 ng/μL) for a total reaction volume of 25 μL. Thermal cycling was performed in a GeneAmp PCR System

Table 1
Samples examined in this study.

Self-identified ethnicity	Number of samples		
	POP[a]	F/S[b]	RM[c]
Caucasian	261	199	12[d]
African American	255	190	
Hispanic	139	197	
Asian	2	200	
Total samples	657	786	12
	1443[e]		12
	1455[f]		

[a] U.S. population samples [18]
[b] U.S. father/son samples [21]
[c] SRM 2391b [22], K562, ABI 007.
[d] The ethnicity is unknown for these samples.
[e] These samples were used for peak height ratio (PHR), stutter percentage calculations and population variation studies.
[f] These samples were used for concordance testing.

[1] Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Table 2
Summary of the concordance comparisons conducted in this study.

Locus	Kit/assay				
	1	2	3	4	5
Amelogenin	PP-ESX17	PP-ESI17	Identifiler	NIST-23plex	MiniFiler
D1S1656	PP-ESX17	PP-ESI17	*a*	*a*	*a*
D2S441	PP-ESX17	PP-ESI17	NIST-23plex	NIST-NC02	*a*
D2S1338	PP-ESX17	PP-ESI17	Identifiler	MiniFiler	*a*
D3S1358	PP-ESX17	PP-ESI17	Identifiler	PP16	*a*
D8S1179	PP-ESX17	PP-ESI17	Identifiler	PP16	*a*
D10S1248	PP-ESX17	PP-ESI17	NIST-23plex	NIST-NC01	*a*
D12S391	PP-ESX17	PP-ESI17	*a*	*a*	*a*
D16S539	PP-ESX17	PP-ESI17	Identifiler	PP16	MiniFiler
D18S51	PP-ESX17	PP-ESI17	Identifiler	PP16	MiniFiler
D21S11	PP-ESX17	PP-ESI17	Identifiler	PP16	MiniFiler
D19S433	PP-ESX17	PP-ESI17	Identifiler	*a*	*a*
D22S1045	PP-ESX17	PP-ESI17	NIST-23plex	NIST-NC01	*a*
FGA	PP-ESX17	PP-ESI17	Identifiler	PP16	MiniFiler
SE33	PP-ESX17	PP-ESI17	PP-SE33	*a*	*a*
TH01	PP-ESX17	PP-ESI17	Identifiler	PP16	*a*
vWA	PP-ESX17	PP-ESI17	Identifiler	PP16	*a*

a—Fewer than five of the eight different kits/assays evaluated in this study provide results for this locus.

9700 (Applied Biosystems) operating in the 9600 emulation mode with the following cycling parameters: a 2-min incubation at 96 °C; followed by 30 cycles of 30 s at 94 °C, 2 min at 60 °C (PP-ESX17) or 59 °C (PP-ESI17), and 90 s at 72 °C; and concluded with a 45-min incubation at 60 °C. A final hold at 4 °C was added until samples were removed.

With the exception of using half reaction volumes, manufacturer recommended conditions were used with the Identifiler, PP16 and PowerPlex® ES SE33 Monoplex (PP-SE33) kits. Manufacturer recommended conditions were used with MiniFiler. Published conditions were used with the NIST-23plex, -NC01, and -NC02 in-house assays.

Following PCR amplification, 1 μL of each sample was diluted in 10 μL Hi-Di formamide (Applied Biosystems) and 1 μL ILS-500 Orange internal size standard (Promega) and analyzed with an ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) using Data Collection v3.0, POP-4 or POP-6 polymer, and a 36-cm array.

2.3. Data analysis

All genotyping was performed with GeneMapper *ID*® v3.2 software (Applied Biosystems) using manufacturer provided allelic ladders, bins, and panels. Once the genotypes were reviewed and edited, data tables were exported into Excel (Microsoft, Redmond, WA). The resultant datasets were compared to determine if there were any discordant allele calls between the PP-ESX17 and -ESI17 Systems. These datasets were also compared to Identifiler, PP16, MiniFiler, PP-SE33, NIST-23plex, NIST-NC01 and NIST-NC02 to identify any further discordance with these PCR amplification kits. A summary of the concordance comparisons conducted for each locus between all datasets is listed in Table 2. Eight different kits/assays were compared, with up to five for a specific locus. Peak height ratios (PHR) between sister alleles, stutter percentages, allele and genotype frequencies, population statistics, and probability of identity (P_I) calculations were performed with Excel based software developed at NIST. These programs are available through the STRBase website: http://www.cstl.nist.gov/biotech/strbase/software.htm. Population statistics calculations including heterozygosities and polymorphism information content (PIC) were confirmed using the PowerMarker v3.25 statistics program [23].

2.4. DNA sequencing

DNA sequencing of all discordant alleles was performed by first amplifying the target sequences for 35 cycles of PCR with the locus specific primers (listed in Supplemental Table S1). The PCR products were run on 32 cm × 0.4 mm polyacrylamide gels and the desired allele bands were excised from the gel and placed overnight in 50 μL of TE^{-4} buffer (10 mmol/L Tris HCl, 0.1 mmol/L EDTA, pH 8.0). A 5 μL aliquot was reamplified for 35 cycles, purified by using ExoSap-IT™ (USB Corporation, Cleveland, OH), and sequenced with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using the unlabeled locus specific primers following the recommendations of the manufacturers. Unincorporated dye terminators were removed using Performa® DTR gel filtration cartridges (Edge Biosystems, Gaithersburg, MD). Samples were analyzed on an ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) using POP-7™ polymer on an 80 cm capillary array. The sequences were aligned using the SeqMan sequencing alignment program (Lasergene, DNA Star, Madison, WI).

3. Results and discussion

3.1. Amplicon size differences between STR kits

The amplicon size differences in base pairs (bp) for each locus in the PowerPlex European Systems were determined for PP-ESI17, PP16, and Identifiler relative to the PP-ESX17 System. The size differences are summarized in Table 3. Several of the primer sets that were used in the new European kits are the same as in PP16. Many of the loci that are smallest in size for PP-ESX17 (<200 bp) were designed to be larger in the PP-ESI17 kit (>200 bp). Since some loci are not present in one or both of the earlier kits (e.g. D1S1656 and D12S391), not all comparisons are possible.

3.2. Concordance evaluation

Concordance evaluations for PP-ESX17 and -ESI17 were performed by comparing the two sets of typing results with each other and with other available multiple PCR amplification kits. Since these multiplex kits are composed of different primer sequences, such concordance checks can identify incorrect or null alleles present in a dataset because of primer binding site mutations. A summary of all of the discordant allele calls found in each kit for all loci tested is listed in Table 4. For each dataset, the number of samples tested varied for each kit as well as the total number of alleles compared. The number of alleles compared was calculated by multiplying the number of samples tested with the number of common loci between kits. This result is doubled

Table 3
STR loci amplicon size differences for three kits relative to the PowerPlex ESX 17 System.

Locus	Size differences relative to PP-ESX17, bp[a]		
	PP-ESI17	PP16	Identifiler
Amelogenin X	0	19	19
Amelogenin Y	0	19	19
D1S1656	87	*b*	*b*
D2S441	260	*b*	*b*
D2S1338	26	*b*	74
D3S1358	0	0	−1
D8S1179	−128	0	−78
D10S1248	200	*b*	*b*
D12S391	164	*b*	*b*
D16S539	−187	−11	−42
D18S51	−151	0	−19
D19S433	−30	*b*	−100
D21S11	0	0	−17
D22S1045	225	*b*	*b*
FGA	−118	55	-56
SE33	42	*b*	*b*
TH01	-85	0	11
vWA	0	0	28

a—Negative values indicate that the PCR product for the compared kit is smaller than in PP-ESX17, positive values indicate that the product for the compared kit is larger than in PP-ESX17, zero indicates that the products are of the same size.
b—The compared kit does not provide results for this locus.

because of the two possible allele combinations for each locus (including homozygotes). The concordance rate is then calculated by dividing the total number of concordant (matching) alleles by the number of potential alleles tested. All of these results are summarized in Table 5. The concordance rates for each dataset compared were greater than 99.79% for the PP-ESX17 System and 99.83% for the PP-ESI17 System.

The majority of the discordant results in this dataset were due to null alleles in which one allele was missing in the profile. All of the primer binding site mutations were determined and confirmed by DNA sequencing. However, there was one discordant result due to a sequence difference outside of the primer positions for the new kits with the SE33 locus. A three base pair deletion (TTG) was found 41 bp downstream from the repeat, thus causing the sample at SE33 to type as a 29.2 or 28.3 depending on the primer set used.

The 10 genomic components of SRM 2391b were typed with the PP-ESX17 and -ESI17 kits as well as two common positive controls, K562 and ABI 007. The results from both kits were in 100% concordance with the published values for these materials.

The final kit configurations were slightly different than the beta-test materials due to additional D22S1045 forward primer in the PP-ESX17 kit and a redesign for the D1S1656 reverse primer in the PP-ESI17 kit. The four discordant samples between kits for the D22S1045 were retested with the final kit and no longer exhibited

Table 4
Discordant results for samples in this study. Null or different alleles due to an insertion or deletion outside of the primer binding site are in bold and underlined.

Locus	PP-ESX17	PP-ESI17	Identifiler	PP16	MiniFiler	NIST-NC01	NIST-23plex	PP-SE33
D1S1656	**15.3**, 15.3	14, 15.3	*a*	*a*	*a*	*a*	*a*	*a*
D3S1358	14, 17	14, 17	14, **14**	14, 17	*a*	*a*	*a*	*a*
D16S539	**12**, 12	12, 13	12, 13	12, 13	12, 13	*a*	*a*	*a*
D18S51	13, 15	13, 15	**15**, 15	13, 15	13, 15	*a*	*a*	*a*
D19S433	13, 14	13, 14	**14**, 14	*a*	*a*	*a*	*a*	*a*
D19S433	13, 14.2	13, 14.2	**14.2**, 14.2	*a*	*a*	*a*	*a*	*a*
D22S1045[b]	**17**, 17	15, 17	*a*	*a*	*a*	15, 17	15, 17	*a*
D22S1045[b]	**17**, 17	15, 17	*a*	*a*	*a*	15,17	15,17	*a*
D22S1045[b]	**17**, 17	15, 17	*a*	*a*	*a*	15,17	15,17	*a*
D22S1045[b]	**16**, 16	15, 16	*a*	*a*	*a*	15,16	15,16	*a*
SE33	26.2, 27.2	26.2, 27.2	*a*	*a*	*a*	*a*	*a*	26.2, **26.2**
SE33	20, 28.3	20, 28.3	*a*	*a*	*a*	*a*	*a*	20, **29.2**
SE33	24.2, 28.2	24.2, 28.2	*a*	*a*	*a*	*a*	*a*	**28.2**, 28.2
SE33	21.2, 26.2	21.2, 26.2	*a*	*a*	*a*	*a*	*a*	21.2, **21.2**
SE33	24.2, 25.2	24.2, 25.2	*a*	*a*	*a*	*a*	*a*	24.2, **24.2**
SE33	19, **19**	19, 25.2	*a*	*a*	*a*	*a*	*a*	19, 25.2

a—The compared kit does not provide results for this locus.
b—After inclusion of an additional D22S1045 forward primer to correct the null allele, these samples are not discordant in the commercial PP-ESX17 kit.

Table 5
Concordance results for multiple kit comparisons.

Kits compared	Number of					
	Common loci	Samples tested	Alleles tested[b]	Alleles match[c]	Discordant samples[d]	Concordance rate (%)[e]
Prototype PP-ESX17 vs. PP-ESI17	17	1443	49,062	49,055	7	99.986
Final PP-ESX17[a] vs. PP-ESI17	17	1443	49,062	49,059	3	99.994
PP-ESX17 vs. Identifiler	11	1445	31,790	31,785	5	99.984
PP-ESI17 vs. Identifiler	11	1445	31,790	31,786	4	99.987
PP-ESX17 vs. PP16	9	659	11,862	11,861	1	99.992
PP-ESI17 vs. PP16	9	659	11,862	11,862	0	100
PP-ESX17 vs. PP-SE33	1	1457	2914	2908	6	99.794
PP-ESI17 vs. PP-SE33	1	1449	2898	2893	5	99.827
PP-ESX17 vs. MiniFiler	6	1123	13,476	13,475	1	99.993
PP-ESI17 vs. MiniFiler	6	1123	13,476	13,476	0	100
Final PP-ESX17[a] vs. NIST-23plex	3	1452	8712	8712	0	100
PP-ESI17 vs. NIST-23plex	3	1452	8712	8712	0	100
Final PP-ESX17[a] vs. NIST-NC01 and NC02	3	663	3978	3978	0	100
PP-ESI17 vs. NIST-NC01 and NC02	3	663	3978	3978	0	100

[a] Includes additional D22S1045 primer.
[b] 2 × (number of samples tested) × (number of common loci) because there are two potential alleles at each locus (including homozygotes).
[c] Concordance as compared to the number of alleles tested.
[d] Samples with null alleles.
[e] 100 × (number of alleles match)/(number of alleles tested).

Table 6
Empirical stutter percentiles for each locus of the PowerPlex ESX 17 and ESI 17 Systems. The data presented represent the stutter percentages where the data fall below the percentiles listed (e.g., 100% of the data at the TH01 locus is below 5.2% stutter). Allele-specific stutter values are provided in the Supplementary information. The loci for both systems are sorted in order from the smallest median (50%) value (top) to the largest (bottom).

Locus	PP-ESX17				Locus	PP-ESI17			
	50%	90%	95%	100%		50%	90%	95%	100%
TH01	1.8	3.1	3.4	5.2	TH01	1.7	3.2	3.5	5.5
D2S441	4.8	6.7	7.2	9.8	D2S441	4.2	6.1	6.5	9.6
D16S539	5.8	8.6	9.0	11.1	D16S539	5.3	7.5	7.9	9.3
D8S1179	6.1	7.9	8.7	15.8	FGA	5.7	8.5	9.1	24.5
D19S433	6.2	8.2	8.8	11.4	D19S433	5.7	8.5	9.3	13.6
FGA	6.5	9.1	9.7	11.9	D8S1179	6.0	8.0	8.6	11.0
D18S51	6.9	9.8	10.4	13.3	D10S1248	6.7	8.6	9.2	11.8
D21S11	7.0	8.9	9.5	15.5	vWA	6.8	8.8	9.4	15.6
vWA	7.1	9.3	9.7	14.4	SE33	6.9	9.4	10.1	18.7
D2S1338	7.3	9.4	10.0	13.2	D1S1656	6.9	10.2	11.3	16.2
D3S1358	8.2	10.0	10.5	12.3	D21S11	6.9	9.8	11.6	20.5
D10S1248	8.2	10.4	11.0	16.4	D18S51	7.4	10.4	11.3	24.0
D12S391	8.2	11.9	13.0	16.1	D12S391	7.7	12.1	13.3	28.9
SE33	8.3	10.6	11.5	15.2	D2S1338	8.0	10.4	11.5	17.2
D1S1656	8.5	11.5	12.3	18.1	D3S1358	8.1	9.9	10.4	13.9
D22S1045	8.9	14.5	16.5	21.0	D22S1045	9.3	15.8	16.9	24.7

the null allele when using PP-ESX17. In addition, a total of 190 samples were run with the final PP-ESX17 kit to determine concordance for the D1S1656 locus with the new reverse primer. All results were concordant, indicating that the performances of both kits are optimal in their final configurations.

3.3. Mutation rate study

Mutation rates for the loci in the PP-ESX17 and -ESI17 Systems were analyzed using the 393 father/son sample pairs. Seven autosomal mutations were observed in four Hispanic, two African American, and one Asian sample pairs. No autosomal mutations were seen in the Caucasian samples. No multiple mutations were observed for the same father/son pair. One mutation was observed in three loci (D3S1358, D8S1179, and D21S11) and two mutations each were observed in the D12S391 and SE33 loci. These results are summarized in Supplemental Table S2. Additional data are needed to more conclusively estimate mutation rates for the newer loci, especially D1S1656 and D12S391. It is important to note that the mutation rates calculated in this study reflect those from paternal origin only and do not account for potential maternal mutations.

3.4. Allele frequency calculations

Allele frequencies were calculated for all of the samples and then separately for four populations (African American, Caucasian, Hispanic and Asian). All allele frequency data from this study can be found in Supplemental Table S3 and on STRBase [17].

3.5. Peak height ratio calculations

The peak height ratios (PHRs) were calculated for all loci in the PP-ESX17 and -ESI17 Systems. Calculations excluded peak heights below 50 Relative Fluorescent Units (RFUs) and above 5000 RFUs. Data below 50 RFUs cannot be considered reliable as compared to the background noise of the instrument. Data above 5000 RFUs was excluded because peak heights above this limit are typically off-scale which causes bleed-through to other dye channels, unusually high stutter peaks, adenylation issues, and locus-to-locus imbalance.

PHRs were determined for sister alleles at every locus by dividing the RFU for the allele with the lower peak height by the RFU for the allele with the higher peak height. The PHR mean was calculated for each genotype combination and then averaged for each locus. For the PP-ESX17 System, the median PHRs were between 82.3% (SE33) and 88.1% (TH01). For the PP-ESI17 System, the median PHRs were between 77.2% (D2S1338) and 90.4% (TH01). The average PHRs were less than 80% only for the four loci: D2S1338, D22S1045, D2S441, and SE33. The complete PHR datasets for each kit can be found in Supplemental Tables S4 (PP-ESX17) and S5 (PP-ESI17).

3.6. Stutter percentage calculations

Stutter products are amplicons that are one repeat less in size than the true allele [13]. Stutter percentage was calculated by dividing the peak height of the stutter product by the peak height of the true allele. Calculations excluded peak heights below 50 RFUs and above 5000 RFUs. Stutter tends to increase with the number of contiguous repeats. Supplemental Tables S6 (PP-ESX17) and S7 (PP-ESI17) list robust estimates for the characteristic value and spread of stutter for all of the frequently observed alleles. Summary values for the stutter at each locus were estimated from the empirical distribution of stutter over all the alleles of the locus. Table 6 lists the observed 50%, 90%, and 95% stutter thresholds and the maximum (100%) stutter values for PP-ESX17 and -ESI17.

The TH01 locus had the lowest 95% stutter threshold in both kits (3.4% and 3.5%, respectively) and D22S1045 had the highest 95% (16.5% and 16.9%, respectively). D22S1045 is a trinucleotide repeat, so it is expected that this locus would possess a higher stutter threshold percentage compared to tetranucleotide repeat loci. The general trend of increasing stutter with increasing number of repeats is particularly pronounced with D22S1045 (Table 7).

3.7. Population variation

Population statistics including observed heterozygosities (H_{obs}), polymorphism information content (PIC) values, and probability of identity (P_I) values were calculated as described previously [13] with the dataset of 1443 samples (Table 8). The number of alleles and the number of observed genotypes at each locus are also listed. The population statistics are the same for both kits because the allele calls were almost identical (only three true discordancies). The loci in Table 8 are sorted in order from the smallest H_{obs} value (top) to largest (bottom). TH01 has the lowest H_{obs} value at 0.7484 and SE33 has the highest value at 0.9383. The complete set of genotype data for both kits is listed in Supplemental Table S8.

Table 7
Allele-specific stutter for D22S1045. As the allele size gets larger, the median stutter percentages increase.

PP-ESX17			Stutter			PP-ESI17			Stutter		
Locus	Allele	Size	#	Median	SD[a]	Locus	Allele	Size	#	Median	SD[a]
D22S1045	10	84.5	20	1.8	0.2	D22S1045	10	308.8	20	1.9	0.4
	11	87.4	132	3.0	0.4		11	311.8	98	2.8	0.5
	12	90.4	36	4.2	0.4		12	314.8	32	4.5	0.6
	14	96.4	48	7.0	0.6		14	321.0	29	6.3	0.6
	15	99.4	167	8.9	0.6		15	323.9	143	9.8	1.6
	16	102.4	120	10.5	1.0		16	327.1	96	9.7	0.9
	17	105.5	105	14.7	3.7		17	330.1	95	14.2	3.0
		Total	628					Total	513		
		Avg		7.2	1.5			Avg		7.0	1.4
		SD		4.6				SD		4.4	

[a] Standard deviation of the median.

Table 8
Population statistics estimated from the NIST U.S. population and U.S. father/son samples for each locus in the PP-ESX17 and PP-ESI17 Systems.

Locus	Unique alleles	Unique genotypes	H (obs)[a]	PIC[b]	P_I[c]
TH01	8	25	0.7491	0.7568	0.0757
D3S1358	11	31	0.7498	0.7304	0.0907
D22S1045	11	45	0.7561	0.7308	0.0933
D2S441	15	47	0.7769	0.7490	0.0805
D16S539	9	30	0.7796	0.7646	0.0724
D10S1248	12	41	0.7810	0.7455	0.0829
D8S1179	11	48	0.7976	0.7960	0.0556
vWA	11	42	0.7997	0.7861	0.0620
D19S433	16	83	0.8094	0.7982	0.0537
D21S11	28	95	0.8302	0.8290	0.0401
D12S391	24	120	0.8663	0.8649	0.0279
FGA	29	111	0.8697	0.8596	0.0300
D18S51	23	103	0.8711	0.8697	0.0263
D2S1338	13	73	0.8732	0.8818	0.0221
D1S1656	17	100	0.8863	0.8805	0.0230
SE33	58	343	0.9383	0.9425	0.0062

[a] Observed heterozygosity.
[b] Polymorphism information content.
[c] Probability of identity.

Table 9
Probability of identity with various sets of STR loci as estimated using population statistics derived from the NIST U.S. population and U.S. father/son samples.

Set	Number of STR Loci	Probability of identity (P_I)				
		Total	Caucasian	African American	Hispanic	Asian
ESS	7	7.5×10^{-10}	1.8×10^{-9}	1.1×10^{-9}	1.1×10^{-9}	8.1×10^{-9}
SGM Plus	10	6.5×10^{-14}	3.9×10^{-13}	8.8×10^{-14}	2.3×10^{-13}	1.7×10^{-12}
CODIS 13	13	4.6×10^{-16}	3.0×10^{-15}	8.9×10^{-16}	1.5×10^{-15}	8.2×10^{-15}
PP16[a]	15	2.9×10^{-19}	5.9×10^{-18}	8.9×10^{-19}	*b*	*b*
Identifiler	15	5.5×10^{-19}	6.9×10^{-18}	7.0×10^{-19}	2.9×10^{-18}	1.9×10^{-17}
PP-ESX16 and -ESI16	15	2.6×10^{-20}	2.0×10^{-19}	3.9×10^{-20}	4.5×10^{-19}	3.1×10^{-18}
PP-ESX17 and -ESX17	16	1.6×10^{-22}	1.4×10^{-21}	4.0×10^{-22}	3.9×10^{-21}	3.6×10^{-20}

a—calculated using just the 657 NIST U.S. population samples.
b—Insufficient samples (<200 samples) for this analysis.

The P_I value for the combined loci for the PP-ESX17 and -ESI17 Systems is 1.61×10^{-22} (Table 9). When the SE33 locus is removed from the calculation (i.e., PP-ESX16 and PP-ESI16), the combined P_I value decreases to 2.58×10^{-20}. Thus, the addition of the highly polymorphic locus SE33 significantly helps [24]. The P_I values were also calculated for additional commercial kits including Identifiler, PP16 and SGM Plus as well as the seven original ESS loci and the 13 Combined DNA Index System (CODIS) loci (Table 9).

The P_I values were determined for each population in this dataset (Table 9). With the exception of the original ESS loci, discriminatory power is greatest for the African American and smallest for the Asian populations. The P_I values for the Caucasian and Hispanic samples were comparable for all of the loci groupings evaluated.

4. Conclusions

The new PP-ESX17 and -ESI17 Systems produce well-balanced PCR products across all loci tested and offer high powers of discrimination with the extended European Standard Set. Both kits have high concordance rates across all NIST samples tested and enable robust comparisons to current commercial multiplex kits including Identifiler and PP16.

While these kits offer high powers of discrimination as compared to the other PCR amplification kits, two of the loci (vWA and D12S391) included in both kits are located in close proximity, approximately 6.36 Mb apart, on chromosome 12. Further studies are planned to examine if these loci are

independently inherited and thus can be appropriately combined using the product rule. The D2S441 and D2S1338 loci are both located on chromosome 2; however, they are positioned on opposite ends of chromosome 2 (150.5 Mb apart) and can be considered genetically unlinked.

The NIST team is developing new software tools to enable rapid concordance data evaluation, allele sharing for parentage samples, allele and genotype frequency calculations, and other pertinent parameters such as heterozygote peak height ratio, inter-locus balance, stutter quantity measurements, population statistics and probability of identity values. As these tools become available, they will be released free-of-charge through the NIST STRBase website: http://www.cstl.nist.gov/biotech/strbase/software.htm.

Acknowledgements

The authors express appreciation for the technical assistance of Jan Redman and Richard Schoske in the initial preparation and quantitation of the NIST U.S. population samples used in this study and to Tom Reid for supplying the DNA Diagnostics Center (DDC) father/son samples. This work was funded in part by the National Institute of Justice (NIJ) through an interagency agreement 2008-DN-R-121 with the NIST Office of Law Enforcement Standards. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2010.03.014.

References

[1] B.S. Weir, Matching and partially-matching DNA profiles, J. Forensic Sci. 49 (2004) 1009–1014.
[2] B.S. Weir, The rarity of DNA profiles, Ann. Appl. Stat. 1 (2007) 358–370.
[3] P.M. Schneider, Expansion of the European Standard Set of DNA database loci—the current situation, Profiles DNA 12 (2009) 6–7.
[4] ENFSI document on DNA-database management April 2009: http://www.enfsi.eu/page.php?uid=98. (accessed December 4, 2009).
[5] P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases—recommendations for new European loci, Forensic Sci. Int. 156 (2006) 242–244.
[6] P. Gill, L. Fereday, N. Morling, P.M. Schneider, Letter to the Editor: new multiplexes for Europe–Amendments and clarification of strategic development, Forensic Sci. Int. 163 (2006) 155–157.
[7] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43–53.
[8] A. Carracedo, M.V. Lareu, Development of new STRs for forensic casework: criteria for selection, sequencing & population data and forensic validation, Proceedings of the 9th International Symposium on Human Identification, available at http://www.promega.com/geneticidproc/ussymp9proc/content/21.pdf. (accessed December 4, 2009).
[9] C.J. Sprecher, R.S. McLaren, D. Rabbach, B. Krenke, M.G. Ensenberger, P.M. Fulmer, L. Downey, E. McCombs, D.R. Storts, PowerPlex ESX and ESI Systems: a suite of new STR systems designed to meet the changing needs of the DNA-typing community, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 2–4.
[10] Technical Manual for PowerPlex® ESX 17 System: http://www.promega.com/tbs/tmd024/tmd024.pdf. (accessed December 4, 2009).
[11] Technical Manual for PowerPlex® ESI 17 System: http://www.promega.com/tbs/tmd028/tmd028.pdf. (accessed December 4, 2009).
[12] STRBase null allele website: http://www.cstl.nist.gov/biotech/strbase/NullAlleles.htm. (accessed December 4, 2009).
[13] J.M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, second ed., Elsevier Academic Press, 2005.
[14] J.M. Butler, Genetics and genomics of core STR loci used in human identity testing, J. Forensic Sci. 51 (2006) 253–265.
[15] M.C. Kline, B. Jenkins, S. Rodgers, Non-amplification of a vWA allele, J. Forensic Sci. 43 (2003) 250–251.
[16] B. Leclair, C.J. Fregeau, K.L. Bowen, R.M. Fourney, Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples, J. Forensic Sci. 49 (2004) 968–980.
[17] NIST population data website: http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm. (accessed December 4, 2009).
[18] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations, J. Forensic Sci. 48 (2003) 908–911.
[19] C.R. Hill, M.C. Kline, J.J. Mulero, R.E. Lagace, C.-W. Chang, L.K. Hennessy, J.M. Butler, Concordance study between the AmpFlSTR MiniFiler PCR amplification kit and conventional STR typing kits, J. Forensic Sci. 52 (2007) 870–873.
[20] J.M. Butler, C.R. Hill, A.E. Decker, M.C. Kline, T.M. Reid, P.M. Vallone, New autosomal and Y-chromosome STR loci: characterization and potential uses, Proceedings of the 18th International Symposium on Human Identification, available at http://www.promega.com/geneticidproc/ussymp18proc/oralpresentations/Butler.pdf. (accessed December 4, 2009).
[21] A.E. Decker, M.C. Kline, J.W. Redman, T.M. Reid, J.M. Butler, Analysis of mutations in father-son pairs with 17 Y-STR loci, Forensic Sci. Int.: Genet. 2 (2008) e31–e35.
[22] NIST SRM 2391b PCR-based DNA Profiling Standard: https://www-s.nist.gov/srmors/view_detail.cfm?srm=2391B. (accessed December 4, 2009).
[23] J. Liu, S.V. Muse, PowerMarker: integrated analysis environment for genetic marker data, Bioinformatics 21 (2005) 2128–2129, Available for download: http://statgen.ncsu.edu/powermarker/index.html. (accessed December 4, 2009).
[24] J.M. Butler, C.R. Hill, M.C. Kline, D.L. Duewer, C.J. Sprecher, R.S. McLaren, D.R. Rabbach, B.E. Krenke, D.R. Storts, The single most polymorphic STR locus: SE33 performance in U.S. populations, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 23–24.